R Notebook

```r
#Ex1
data1<-read.table("/Users/phamkhoa/Documents/university/3/stats_model_1/w3/galapagos.txt", sep="\t", de
attach(data1)
```

```r
#a
poi.log<-glm(Species~Area+Elevation+Nearest+Scruz+Adjacent, family = poisson(link = "log"), data = data

newdata<-data.frame(Area = 58.27, Elevation= 198, Nearest = 1.1, Scruz = 88.3, Adjacent = 0.57)
eta<-predict(poi.log, newdata=newdata, type="link", se.fit=TRUE)
link.lowerbound<-eta$fit-qnorm(0.975)*eta$se.fit
link.upperbound<-eta$fit+qnorm(0.975)*eta$se.fit

lower<-exp(link.lowerbound)
upper<-exp(link.upperbound)
lower
```

```
##        1
## 25.01812
```

```r
upper
```

```
##        1
## 31.04843
```

```r
#b
H0 <- glm(Species~Area, family = poisson(link = "sqrt"), data = data1)
H1 <- glm(Species~Area+Elevation+Nearest+Scruz+Adjacent, family = poisson(link = "sqrt"), data = data1)
test<- anova(H0, H1, test = "Chi")
test$Deviance[2]
```

```
## [1] 1618.246
```

```r
test$`Pr(>Chi)`[2]
```

```
## [1] 0
```

```r
#Deviance = 1618.246
#Pr(>Chi) = 0
```

```r
#c
model <-glm(Species~log(Area)+log(Elevation)+log(Nearest)+log(Scruz)+log(Adjacent), family = poisson(li
newdata<-data.frame(Area = 58.27, Elevation= 198, Nearest = 1.1, Scruz = 88.3, Adjacent = 0.57)
pred<-predict(model, newdata=newdata, type="response")
```

```
xf<-c(1,58.27, 198, 1.1, 88.3, 0.57)

Var.eYf<-pred*(1+pred*t(xf)%*%vcov(model)%*%xf)
lower.Yf<-pred-qnorm(0.9)*sqrt(Var.eYf)
upper.Yf<-pred+qnorm(0.9)*sqrt(Var.eYf)
lower.Yf
```

```
##            [,1]
## [1,] -1471.833
```

```
upper.Yf
```

```
##           [,1]
## [1,] 1717.442
```

```
#d
library(MASS)
modelNB.log12345<-glm.nb(Species~Area+Elevation+Nearest+Scruz+Adjacent, data = data1)
summary(modelNB.log12345)
```

```
##
## Call:
## glm.nb(formula = Species ~ Area + Elevation + Nearest + Scruz +
##     Adjacent, data = data1, init.theta = 1.674581694, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1344  -0.8598  -0.1476   0.4576   1.8416
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.9064844  0.2510358  11.578  < 2e-16 ***
## Area        -0.0006336  0.0002865  -2.211 0.027004 *
## Elevation    0.0038552  0.0006916   5.574 2.49e-08 ***
## Nearest      0.0028246  0.0136616   0.207 0.836202
## Scruz       -0.0018968  0.0028096  -0.675 0.499604
## Adjacent    -0.0007605  0.0002278  -3.339 0.000842 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.6746) family taken to be 1)
##
##     Null deviance: 88.430  on 29  degrees of freedom
## Residual deviance: 33.196  on 24  degrees of freedom
## AIC: 304.22
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  1.675
##           Std. Err.:  0.442
##
##  2 x log-likelihood:  -290.223
```

```
modelNB.log12345$fitted.values[1]
```

```
##        1
## 68.28233
```

```
#68.28233
```

```
data2<-read.table("/Users/phamkhoa/Documents/university/3/stats_model_1/w3/chromoabnormal.txt", sep="\t
attach(data2)
```

```
#Ex2
#a
m12<-glm(ca~offset(log(cells)) + doseamt*doserate, family = poisson(link = "log"), data = data2)
newdata <-data.frame(doseamt = 4, doserate = 0.75, cells = 64070)
pred<-predict(m12, newdata=newdata, type="response")
pred
```

```
##        1
## 311.3886
```

```
#b
newdata <-data.frame(doseamt = 4, doserate = 0.75, cells = 64070)
pred<-predict(m12, newdata=newdata, type="response")
pred
```

```
##        1
## 311.3886
```

```
ratio.prediction<-pred/newdata$cells
ratio.prediction
```

```
##          1
## 0.004860132
```

```
xf<-t(cbind(1,4,0.75,3))
```

```
Var.eYf<-pred*(1+pred*t(xf)%*%vcov(m12)%*%xf)
```

```
lower.Yf<-pred-qnorm(0.9)*sqrt(Var.eYf)
upper.Yf<-pred+qnorm(0.9)*sqrt(Var.eYf)
lower.Yf
```

```
##           [,1]
## [1,] 286.4762
```

```
upper.Yf
```

```
##           [,1]
## [1,] 336.3011
```

```
Var.eZf<-((1/newdata$cells)^2)*Var.eYf

lower.Zf<-ratio.prediction-qnorm(0.9)*sqrt(Var.eZf)
upper.Zf<-ratio.prediction+qnorm(0.9)*sqrt(Var.eZf)
lower.Zf
```

```
##            [,1]
## [1,] 0.0044713
```

```
upper.Zf
```

```
##              [,1]
## [1,] 0.005248963
```

```
#c
H0.quassi <-glm(ca~offset(log(cells)) + doseamt, family = quasipoisson(link = "log"), data = data2)
#Since we test the significant of X2, I test both B2 and B3
H1.quassi <-glm(ca~offset(log(cells)) + doseamt*doserate, family = quasipoisson(link = "log"), data = d

anova(H0.quassi,H1.quassi, test = "F")
```

```
## Analysis of Deviance Table
##
## Model 1: ca ~ offset(log(cells)) + doseamt
## Model 2: ca ~ offset(log(cells)) + doseamt * doserate
##   Resid. Df Resid. Dev Df Deviance      F  Pr(>F)
## 1        25     409.40
## 2        23     270.26  2   139.14 5.3631 0.01225 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(H0.quassi,H1.quassi, test = "F")$F[2]
```

```
## [1] 5.363088
```

```
#F = 5.363088
anova(H0.quassi,H1.quassi, test = "F")$`Pr(>F)`[2]
```

```
## [1] 0.01225317
```

```
#p = 0.01225317
```

```
#d
i<-glm(ca~offset(log(cells)) + doseamt*doserate, family = poisson(link = "log"), data = data2)
summary(i)
```

```
##
## Call:
## glm(formula = ca ~ offset(log(cells)) + doseamt * doserate, family = poisson(link = "log"),
```

```
##      data = data2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.7308  -2.2842  -0.6264   3.3487   5.8272
##
## Coefficients:
##                  Estimate Std. Error  z value Pr(>|z|)
## (Intercept)      -7.90511    0.06160 -128.322  < 2e-16 ***
## doseamt           0.61224    0.01707   35.862  < 2e-16 ***
## doserate          0.06401    0.02922    2.191 0.028476 *
## doseamt:doserate  0.02715    0.00765    3.549 0.000387 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4753.00  on 26  degrees of freedom
## Residual deviance:  270.26  on 23  degrees of freedom
## AIC: 453.67
##
## Number of Fisher Scoring iterations: 4
```

```
ii<-glm(ca~offset(log(cells)) + doseamt*doserate, family = quasipoisson(link = "log"), data = data2)
summary(ii)
```

```
##
## Call:
## glm(formula = ca ~ offset(log(cells)) + doseamt * doserate, family = quasipoisson(link = "log"),
##      data = data2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.7308  -2.2842  -0.6264   3.3487   5.8272
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -7.90511    0.22188 -35.628  < 2e-16 ***
## doseamt           0.61224    0.06149   9.957 8.29e-10 ***
## doserate          0.06401    0.10524   0.608    0.549
## doseamt:doserate  0.02715    0.02755   0.985    0.335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 12.97228)
##
##     Null deviance: 4753.00  on 26  degrees of freedom
## Residual deviance:  270.26  on 23  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
```

```
iii<-glm.nb(ca~offset(log(cells)) + doseamt*doserate, data = data2)
summary(iii)
```

```
##
## Call:
## glm.nb(formula = ca ~ offset(log(cells)) + doseamt * doserate,
##     data = data2, init.theta = 10.44130185, link = log)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.6740  -0.9467  -0.4614   0.9319   1.7528
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.82300    0.20683 -37.823   <2e-16 ***
## doseamt           0.60844    0.06216   9.789   <2e-16 ***
## doserate          0.05891    0.09890   0.596    0.551
## doseamt:doserate  0.03338    0.02969   1.124    0.261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(10.4413) family taken to be 1)
##
##     Null deviance: 303.957  on 26  degrees of freedom
## Residual deviance:  27.789  on 23  degrees of freedom
## AIC: 278.85
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  10.44
##           Std. Err.:  3.13
##
##  2 x log-likelihood:  -268.852
```

```
AIC(i)
```

```
## [1] 453.6666
```

```
AIC(ii)
```

```
## [1] NA
```

```
AIC(iii)
```

```
## [1] 278.8522
```

```
MSE.M1<-sum((ca-fitted(i, type="response"))^2)/27
MSE.M2<-sum((ca-fitted(ii, type="response"))^2)/27
MSE.M3<-sum((ca-fitted(iii, type="response"))^2)/27
```

```
MSE.M1
```

```
## [1] 970.778
```

MSE.M2

```
## [1] 970.778
```

MSE.M3

```
## [1] 1549.685
```

#chose i poisson log model